HOW ARGENTINA WON THE 2022 FIFA MEN'S WORLD CUP:

A DATA STORY

by

ANIRUDDHA PARTHASARATHY

A master's capstone project submitted to the Graduate Faculty in Data Analysis and Visualization in partial fulfillment of the requirements for the degree of Master of Science, The City University of New York

2025

APPROVAL

How Argentina Won the 2022 FIFA Men's World Cup:
A Data Story

by

Aniruddha Parthasarathy

This manuscript has been read and accepted for the Graduate Faculty in
Data Analysis and Visualization in satisfaction of the capstone project requirement
for the degree of Master of Science.

Approved: March 2025

Professor T. Howard Everson

Professor Matthew K. Gold

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

How Argentina Won the 2022 FIFA Men's World Cup:
A Data Story

by

Aniruddha Parthasarathy

Advisor: Professor Howard T. Everson

This project analyzes Argentina's 2022 FIFA Men's World Cup win using open-source football (soccer) data. The project evaluates the team's performance at a micro-level across three domains: without possessing the ball, possessing the ball and the team's in-game management tactics. A statistical framework, i.e., multiple linear regression modeling, was used to identify the five key defensive actions influencing the Argentinian team's intensity of pressure applied, and visualized by heatmaps and time-segmented plots. More specifically, an Expected Threat (xT) analysis quantified the threat or danger from passes and progressive carries (moving the ball at least 10 meters), revealing that Lionel Messi's involvement, the team's star player, increased average xT gains by ~17%. Temporal plots, annotated with key events such as substitutions and formation shifts, showed how Argentina maintained ball control under opposition pressure. These analyses suggest that selective pressing, multiple quick transitions, high levels of shooting accuracy, tactical flexibility, and some individual brilliance played pivotal roles in the Argentinian's trophy win. The project website is hosted at: https://a-partha.github.io/Masters-Capstone-Project/, with full project code and datasets available at: https://github.com/a-partha/Masters-Capstone-Project. The analyses presented here contribute to the growing field of applied football (soccer) analytics by offering a reproducible case study of tournament-level

performance. It also serves as a blueprint for integrating large-scale data with interactive visualizations to bridge the gap between technical analyses and broader public engagement.

ACKNOWLEDGMENTS

I would like to thank Prof. Howard T. Everson, Prof. Matthew K. Gold, and Mr. Jason Nielsen for their incredible support and understanding throughout the development of this project.

I also extend my gratitude to my family, friends, and mentors for their constant encouragement and belief in me.

CONTENTS

TABLES

FIGURES

DIGITAL MANIFEST

I. **Capstone Whitepaper (PDF)**

A detailed technical report documenting the methodology, analysis, and insights from the project.

II. **Interactive Website (HTML/CSS/JavaScript)**

A scroll-driven narrative experience built using D3.js, Scrollama, and Cursor, hosted on GitHub Pages.
URL: https://a-partha.github.io/Masters-Capstone-Project/

III. **Code and Data Repository (ZIP archive from GitHub)**

Contains all project files including source code, data visualizations, Python analysis notebook (.ipynb), data (.csv, .json), and assets used in the site.
URL: https://github.com/a-partha/Masters-Capstone-Project

IV. **Website Archive (.warc file)**

An archived copy of the website created using the Conifer web archiving service for long-term preservation.

A NOTE ON TECHNICAL REQUIREMENTS

This project includes an interactive, scroll-driven website built using HTML, CSS, and JavaScript, with data visualizations developed using D3.js and transitions powered by Scrollama. All visualizations are rendered using SVG and JSON inputs, and static images are loaded as separate assets from the project repository. The interaction logic is written in JavaScript. Data processing and analyses were conducted using integrated development environments (IDEs) such as Cursor and Google Colab.

The site was designed specifically for desktop browsers and is not optimized for mobile or tablet viewing. For proper functionality, JavaScript must be enabled in the browser. All code, assets, and datasets are stored in a public GitHub repository, and the final website is hosted on GitHub Pages. A .warc file was generated using Conifer for long-term preservation. Additionally, AI tools such as ChatGPT, Claude, Perplexity, YouLearn AI, and Cursor AI were used during various stages of the project for assistance with coding, debugging, and writing support.

To my loving grandmother Smt. Sarojamma S., whose strength continues to inspire me; to Tuffy, our beautiful companion who brought endless joy to our family; to my niece, whose exciting arrival reminds us that our future is always filled with love and hope; and to the neurodivergent community, whose persistence is nothing short of brilliant.

NARRATIVE

Introduction

Argentina's victory at the 2022 FIFA Men's World Cup has been widely recognized for both its emotional resonance and symbolic importance. From a data science perspective, however, the 2022 tournament offered a unique opportunity to examine the convergence of the team's tactical coherence, adaptability, and individual player performance in a high-stakes environment. This capstone project utilizes open-source event data from Hudl StatsBomb—an industry-standard provider of structured football (soccer) data capturing time-stamped on-the-ball actions (e.g., passes, shots, pressures), each annotated with spatial and contextual metadata. The objective of this project is to provide an example of how a data-driven lens can be used to evaluate how the Argentine team created, limited, and managed ability and skill throughout the games in the tournament.

Recognizing that data analysis is often obscured by dense metrics and overly abstract representations, this project adopts a hybrid approach: the final interactive website presents Argentina's tactical journey in a narrative format, accessible to general audiences, while this whitepaper serves as its technical and methodological companion. All analyses were organized across three primary domains:

1. *Without the Ball:* Examining out-of-possession behavior such as pressure distribution (ball and player movement), and pressing frequency and intensity;
2. *With the Ball:* Examining in-possession behavior such as shot creation, and ball progression patterns; and
3. *In-Game Management:* Assessing substitutions and tactical shifts in response to the dynamics of the game, i.e., the game state.

The analysis of each of these domains was supported by custom-built visualizations that were developed using D3.js, a JavaScript library for data-driven visualizations, and Scrollama, a scroll-driven library that enables narrative transitions based on the user's scroll position. Visualizations include pitch-level pressure heatmaps, zone-to-zone Expected Threat (xT) flows (i.e., variations in chances of scoring), and temporal lollipop plots (discrete value timelines) of Argentina's shooting behavior. Data processing and analyses were conducted in Google Colab, a cloud-based IDE for Python. The final interactive site is hosted via GitHub Pages and includes downloadable code, datasets, and this whitepaper. The primary audience for the interactive site includes football coaches, journalists, and curious fans. In contrast, this whitepaper is intended to document the data processing pipeline, design decisions, and statistical methods that shaped the narrative. It serves both as a technical report and as a demonstration of how open-source football data can be leveraged for analyses that are both accessible and robust.

Relationship to Focus Area and Previous Course of Study

This project also serves as a use case for applying skills taught during coursework to a real-world scenario. Statistical techniques such as exploratory data analysis and regression analysis, learned in the *Data Analysis Methods* course; storytelling and web-designing tools like HTML, CSS, and JavaScript, learned in the *Software Design Lab: Creative Computing* course; and dynamic data visualization methods like D3.js, learned in the *Interactive Data Visualization* course — along with choosing the most appropriate visuals to effectively convey the results, as taught in the *Visualization and Design* course — were all developed through the MS program in Data Analysis and Visualization.

Environmental Scan and Literature Review

The recent surge in data availability and advanced analytical techniques has transformed how professional football is understood, managed, and optimized. Previous literature and publicly available datasets provide critical foundations for this project. Hudl StatsBomb [1, 3] enables a structured approach to player tracking and event data analysis, facilitating detailed insight into on-field events and tactical dynamics. Additionally, Devin Pleuler [2] offers a thorough, hands-on “Soccer Analytics Handbook” on GitHub, serving as both a technical and methodological guide for beginners and seasoned analysts alike. On a broader scale, there is a growing body of research examining operational frameworks and data practices in football. For example, Lolli et al. [4] provides an extensive survey of how professional clubs and national federations integrate data analytics into decision-making processes. This study will benefit from incorporating such frameworks, understanding how clubs and teams operationalize data, and adapting these methods to analyze Argentina’s path to victory in the 2022 FIFA World Cup.

From a historical perspective, articles from The New York Times [5, 7] reveal how football clubs in England have progressively embraced analytics, with data departments now forming an essential part of club structures. These studies emphasize the evolving role of data in football, contextualizing the increasing value of analytics in decision-making. Similarly, Daniel Burdeno [6], lead data science instructor at the Flatiron School (an online private coding bootcamp), explores football data analytics, highlighting the educational value of such work and its impact on analytical literacy among fans, professionals, and aspiring data scientists.

In terms of specific football metrics and models, Power et al. [9] emphasize the differentiation of actions (e.g., passes) based on risk and reward, introducing nuanced metrics that move beyond traditional stats. This project will draw upon these methods to classify and evaluate Argentina's passing and other play patterns to assess their role in successful match outcomes. Moreover, Rudd [10] provides a framework using Markov Chains for tactical analysis, offering a structured approach to quantifying offensive actions that can be applied to Argentina's gameplay. The project also adapts the open-source expected threat (xT) model developed by Singh [11], which quantifies the likelihood of scoring from different pitch zones based on possession actions. Through these methods, the project aims to highlight actionable insights for tactical decision-making, further advancing the application of analytics in professional football.

<u>Context</u>

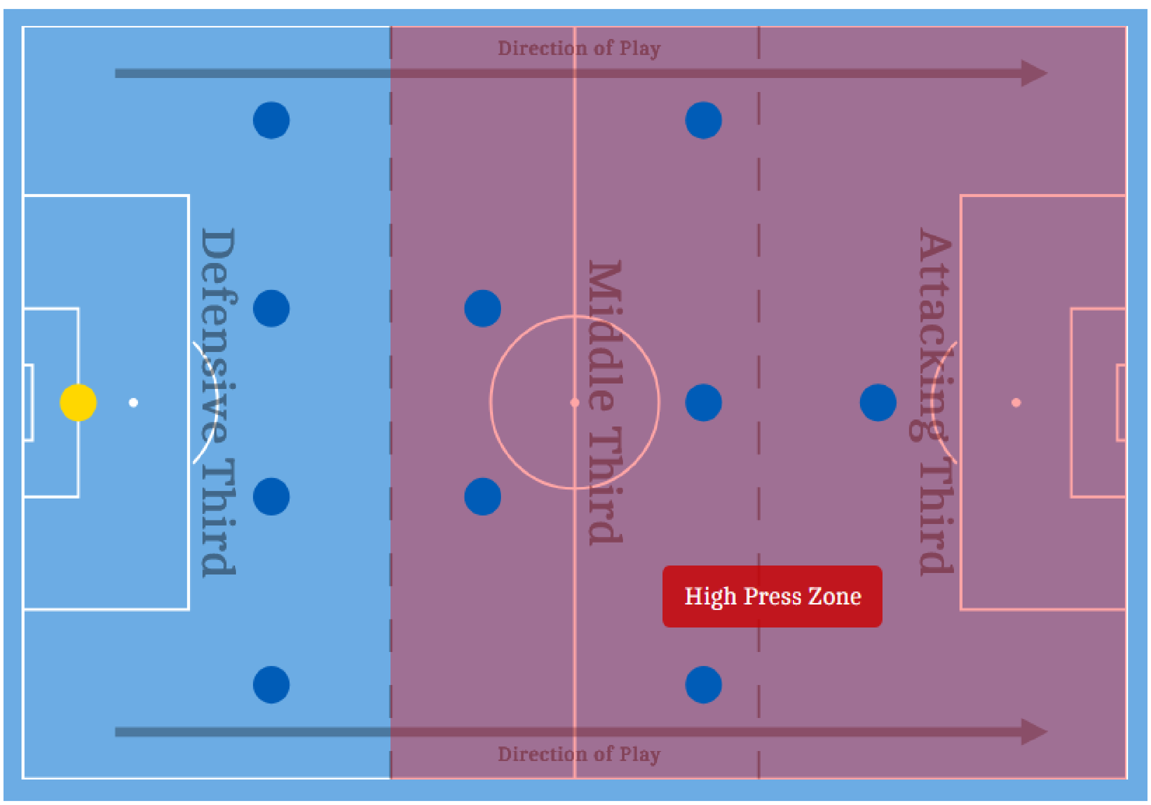

Fig. 1. Example of an Argentine Formation with the High Press Zone.

Figure 1 above is an example of how Argentina's players lined up on average during a game. It is called the 4-2-3-1 formation, but the goalkeeper (shown there in yellow) is not considered in this setup because that position typically remains constant. Argentina also started with or changed to a different formation such as a 4-4-2, 4-3-3, or 3-5-2 depending on whether they were in possession of the ball, or the opponent's formation, or the game state (score), injuries, and other tactical factors. The fewer possessions the opposing teams have, obviously, the fewer chances they have to score. It also means Argentina may have more opportunities to control the ball and create scoring chances themselves. The zone highlighted in red in Figure 1 is the High Press Zone (HPZ) — an area higher up the pitch where Argentina attempted to regain possession of the ball through various defensive actions. The primary objective of engaging in the high pressing tactic is twofold: to recover the ball in advanced zones, thereby increasing the likelihood of immediate attacking opportunities, and to disrupt the opposition's build-up before it could progress toward Argentina's own goal.

<u>Argentina without the Ball</u>

To quantify Argentina's pressing behavior, Passes Per Defensive Action (PPDA) was computed for each match. This metric, widely used in football analytics, measures the total number of passes the opposition completes in the High Press Zone (HPZ) before a defensive action (tackle, interception, block, etc.) is attempted. It is calculated using the formula:

$$PPDA\ =\ (Number\ of\ opposition\ passes\ made)\ /\ (Number\ of\ defensive\ actions\ made)$$

Table. 1. Argentina's PPDA Values and Frequency Counts of Defensive Actions Across All World Cup Games.

Opposition	PPDA	Oppos -ition Passes	50 / 50	Blocks	Tackles	Fouls Comm -itted	Inter cep-tions	Aerial Duels Lost	Aerial Duels Won
Saudi Arabia	10.08	242	5	3	2	6	5	0	3
Mexico	12.00	324	3	2	6	9	2	3	2
Australia	10.13	233	0	2	5	11	4	0	1
Poland	10.38	270	0	1	6	7	8	1	3
Netherlands	20.00	520	2	0	4	14	4	2	0
Croatia	16.30	489	2	0	7	16	4	1	0
France	8.60	473	4	5	12	22	11	1	0

Lower PPDA values correspond to higher pressing intensity. Table 1 above summarizes Argentina's PPDA values across all matches in the tournament, alongside frequency counts of defensive actions contributing to these values. To identify which types of defensive actions most strongly influenced PPDA variation, a multiple linear regression was performed with PPDA as the dependent variable.

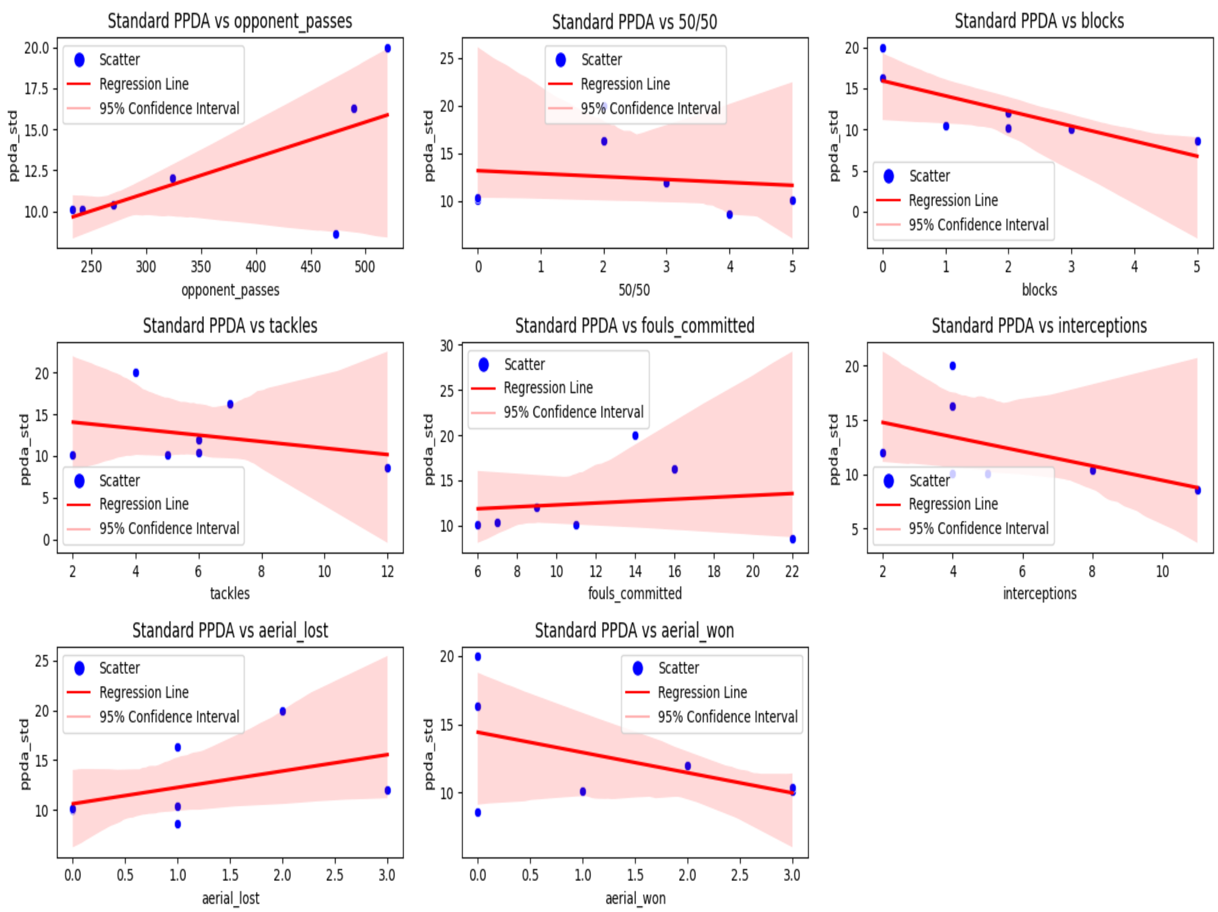

Fig. 2. Scatter Plots with Regression Line and 95% Confidence Interval of Argentina's PPDA vs. Each Contributing Defensive Action per Game.

Each scatter plot in Figure 2 illustrates the marginal relationship between Argentina's PPDA and individual defensive actions across matches, with regression lines and 95% confidence intervals capturing direction and variability. Blocks, interceptions, tackles, and aerial duels won all show negative correlations with PPDA, indicating that more frequent execution of these actions is associated with more intense pressing (lower PPDA). However, more aerial duels lost seem to increase PPDA. But due to multicollinearity among variables, these reflect individual associations rather than isolated causal effects.

Three feature grouping strategies were considered for the independent variables (defensive actions): manually defined tactical groups, correlation-based hierarchical clusters, and a third variant based on the original PPDA definition. Since the actions were highly multi-collinear, centering (mean-adjusting) was used to mitigate it for regression use. The cluster grouped model explained approximately 98% of the variance in PPDA values, with five actions—tackles, blocks, fouls, interceptions, and 50/50s—emerging as the dominant predictors. Empirically, each additional defensive action from this group reduced PPDA by an average of 0.3 passes, reinforcing their functional relevance to Argentina's pressing system.

Building on these statistical findings, subsequent spatial analysis focused on pressure events — defined as moments when an Argentine player actively challenged an opponent receiving, carrying, or releasing the ball without making a tackle or committing a foul. Heatmaps of these events, stratified by match, revealed how pressing concentration evolved over the course of the tournament.

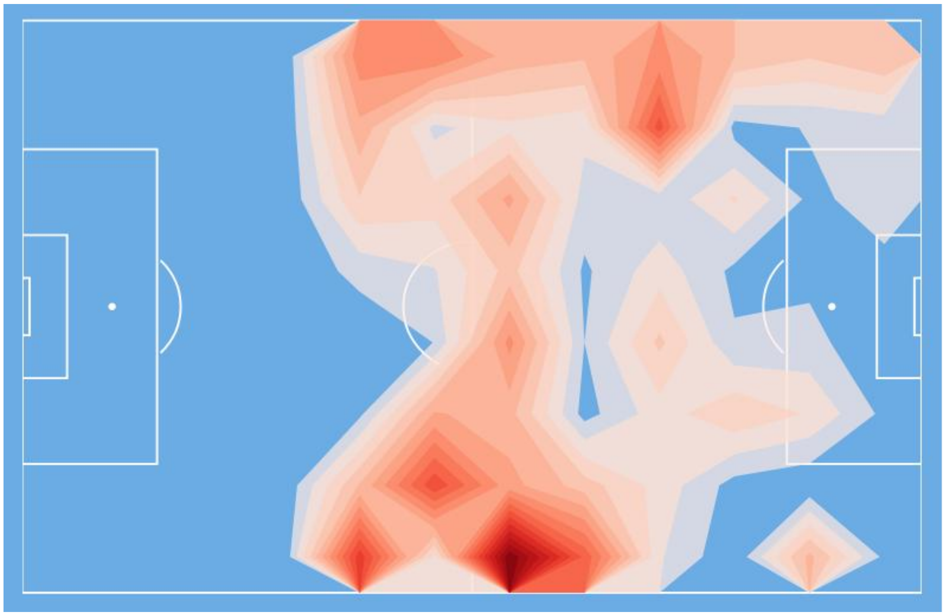

Fig. 3. Argentina's Pressure Heatmap Against France in the Final.

Hotter zones indicate where more pressure events occurred. The heatmaps show how Argentina first addressed their lack of aggression in the midfield, and how they eventually fixed their poor performance in the wide areas while still adapting to opposition ball-playing style. A stacked bar plot was constructed to evaluate the variation of pressure intensity across different periods of normal time (0' to 90') each game.

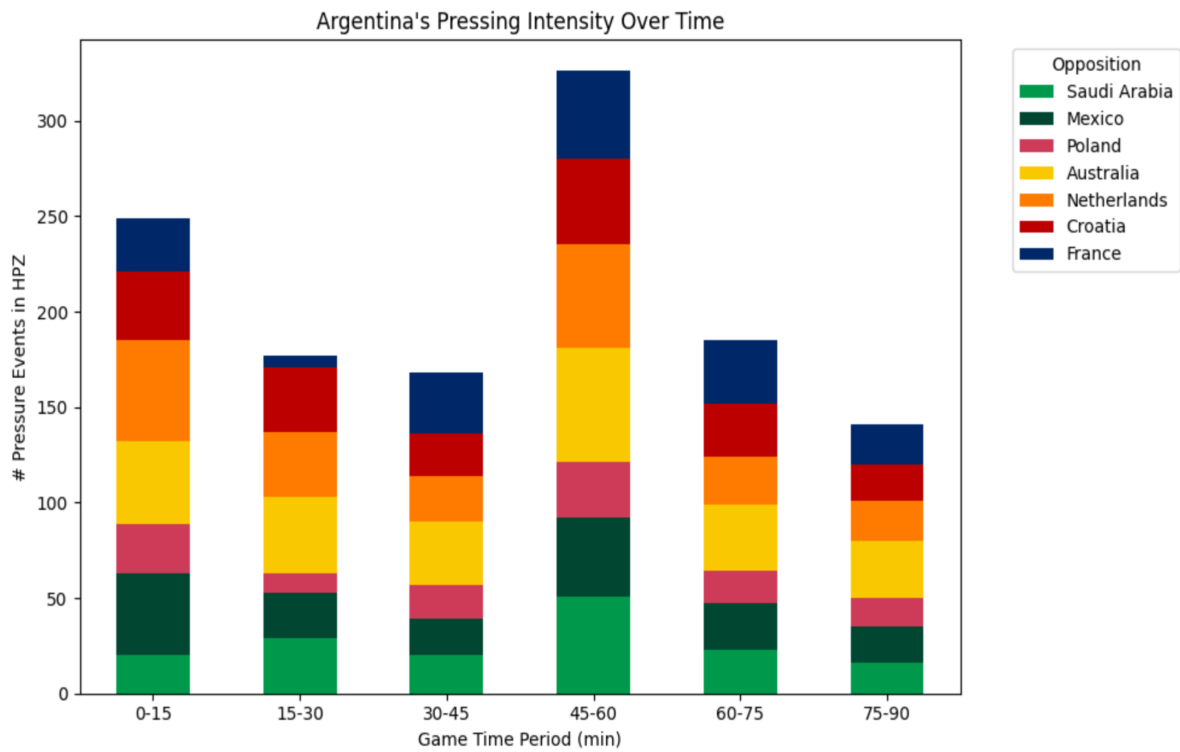

Fig. 4. Argentina's Pressure Events in the HPZ per Game Across 15-Minute Time Intervals.

The plots in Figure 4 reveal a clear temporal trend: Argentina typically applied heightened pressure up front in the first 15 minutes of each half, followed by periods of modulation. In sum, although Argentina's out-of-possession play was not uniformly aggressive, it was strategically deliberate. Pressing was applied selectively, with both spatial and temporal variation, and was consistently followed by subsequent defensive actions.

Argentina with the Ball

When in possession of the ball, Argentina exhibited a structured yet opportunistic approach, marked by a balance between deep build-up sequences and rapid attacking transitions. To evaluate how much high pressing influenced what they did with the ball, the time between high-pressure events and subsequent shot attempts was measured. It was observed that 31.58% of Argentina's 76 total shots were taken within 18.4 seconds from an occurrence of a high-pressure event. These sequences accounted for four goals, representing 28.57% of the team's tournament goal total. A significant proportion of these pressure-induced shots occurred in the early stages of the second half (45' to 60') — a period corresponding to Argentina's highest average pressing intensity. Hence, if Argentina won possession of the ball, they were often quick to take a shot.

Beyond fast transitions, Argentina's build-up play was also impressive. Build-up play refers to sequences involving controlled movement of the ball from the defensive third to the attacking third. This helped identify how Argentina progressed the ball into goal scoring zones. To avoid including plays like direct long balls from defense to attack, only sequences with at least 3 passes or progressive carries (carrying for at least 10 meters) were considered. Across the tournament, 74 such build-up sequences were identified. Most came from wide areas than through the middle — something they improved on with each game.

As noted earlier, a metric called Expected Threat (xT) was used to quantify the danger presented by those sequences. This metric assigns a value to different zones on the pitch based on how likely ball possession in that zone leads to a goal. xT captures not just shots on goal, but also the threat created by passes and carries (where a player moves the ball forward while in control). The expected threat (xT) model used in this study was adapted from an open-source expected threat (xT) model developed by Singh [11], a contributor to football analytics and currently a data scientist for Arsenal Football Club. To keep things as objective as possible, the model was trained only on matches not involving Argentina from the 2022 World Cup. This made it possible to observe how dangerous Argentina’s build-up play would be if they had non-Argentine finishers — as if someone other than the likes of Lautaro Martínez or Lionel Messi were on the end of those chances.

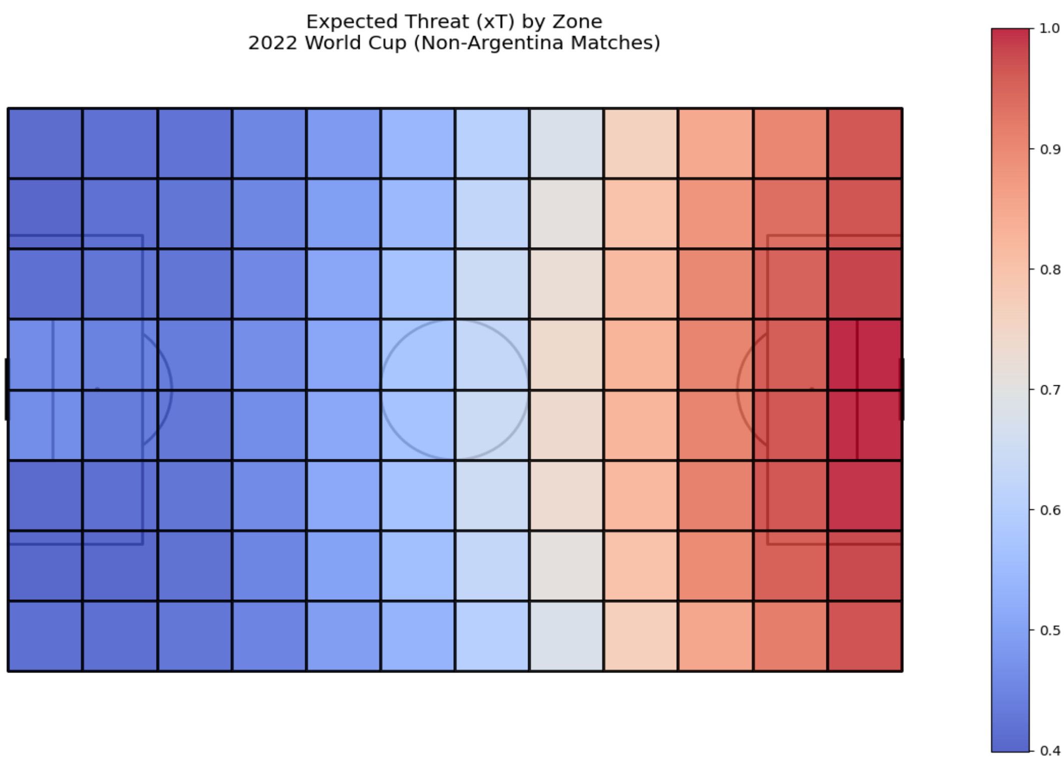

Fig. 5. Heatmap Grid for Average xT of Non-Argentine Teams at the World Cup.

The xT grid above serves as a neutral baseline to measure Argentina's threat from chance creation alone, without factoring in their ability to take those chances (via shots or goals). For example, an xT value of 0.9 means that, on average, a player scored 90% of the time from that zone of the pitch. Naturally, zones in the attacking third closer to the opposition goal pose a greater threat.

xT analysis also helped study the importance of Messi's performances who did not have a fixed position in the games. Sometimes he played higher up the pitch; other times, he dropped deeper to join the build-up. Assessing his influence on build-up play revealed that out of those 72 build-up sequences, Messi was involved in 62. When he was involved in Argentina's build-up play, the average sequence xT gain was 0.51. Without him, the average gain dropped to 0.44. This represents an approximate 17% increase in scoring probability attributable to his involvement in sequence progression. The most impactful sequence with Messi involved yielded a 99% increase in xT, whereas the best sequence without him peaked at a 57% increase. This implies that when Messi helped move the ball forward, Argentina were far more likely to score.

Regarding finishing efficiency, Messi's own shot conversion marginally exceeded expectations: he scored three goals when he was expected to score 2.1 goals (xG) based on the quality of chances he created. The remainder of the squad scored 13 goals from 10.6 xG, reflecting a consistently above-average conversion rate across the team. However, Argentina led the tournament in total shots (139) and non-penalty goals (16), suggesting that a combination of volume and quality contributed to their offensive effectiveness.

Argentina's In-Game Management

This section focuses on how Argentina's coaching team managed to adapt during games. Only games where Argentina conceded at least one goal were considered to show how they responded to adversity. A metric from StatsBomb's event data called "under_pressure" was used for this purpose. It shows how often Argentina moved the ball under opposition pressure per minute. This includes any action (pass, carry, etc.) performed while being pressed anywhere on the pitch. A higher value means Argentina were more comfortable evading pressure by playing through it. Temporal fluctuations along with the average were plotted and annotated with key game events, such as goals, substitutions, and tactical adjustments to help assess its flow.

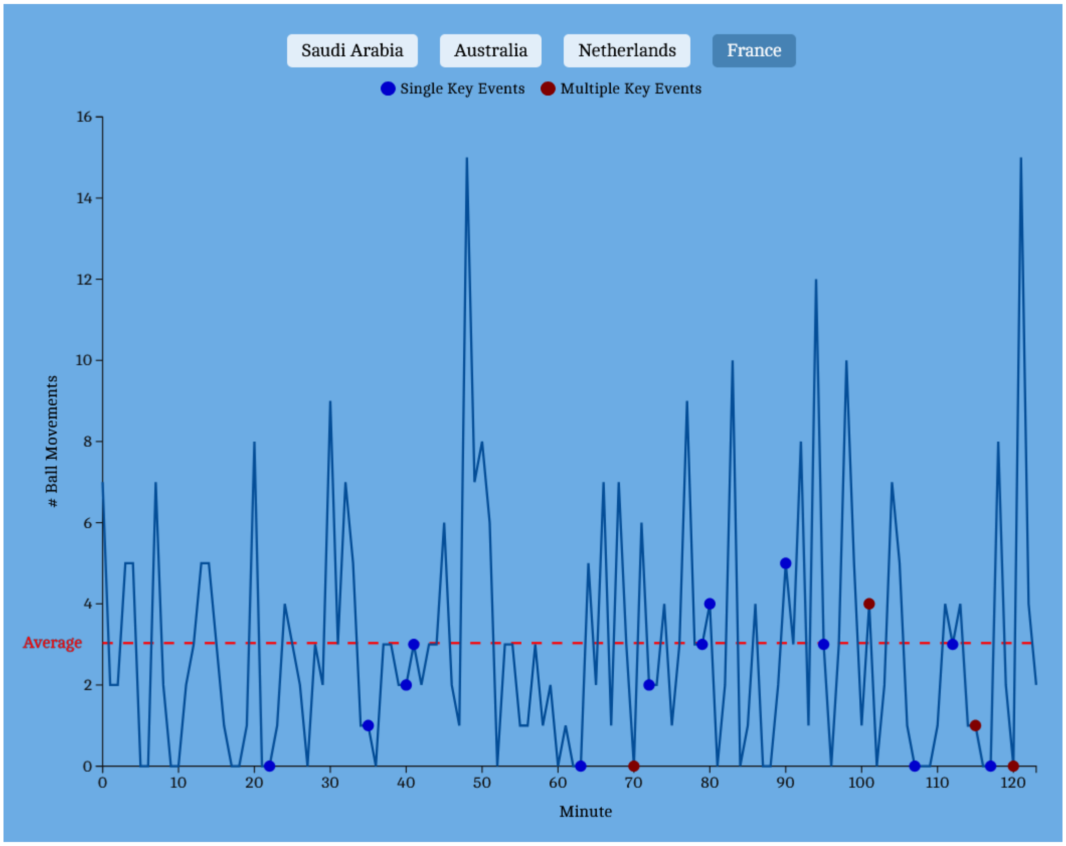

Fig. 6. Number of Times Argentina Moved the Ball per Minute in the Final Against France When Under Pressure.

Following their opening game stumble, Argentina's coaching staff used substitutions and formation changes proactively during games — not just to react, but to stay ahead. But it didn't always work; a switch to a five-man defense formation after taking the lead backfired more than once. However, they always adapted again, and their best control came not from loading the defense, but from staying sharp and balanced in midfield with formations like 4–4–2, 4–2–3–1 and 4–3–3. This adaptability under dynamic match conditions contributed significantly to maintaining possession under pressure, managing the pace of the game, and securing control in key phases of tournament play.

Conclusion

Argentina's 2022 FIFA Men's World Cup campaign offers a compelling case study in how data-informed strategy, tactical flexibility, and individual brilliance combine to produce success in elite tournament football. Through event-level analysis of pressing intensity, ball possession dynamics, and in-game adjustments, this study has highlighted how Argentina balanced proactive and reactive play across different game states. By leveraging open-source data and reproducible methods, this project demonstrates how modern football analytics can identify latent patterns, quantify performance impact, and generate actionable insights for decision makers.

Evaluation and Continuation

This project set out to analyze Argentina's World Cup performance through data-driven methods and present it through an interactive narration. Overall, the completed version strongly captures the intended objectives by providing detailed statistical analyses, clear visual storytelling, and reproducibility through open-source tools. But a few challenges emerged during development. Framing numerical findings into actionable insights that fit into the story was particularly tricky. Finding ways to group defensive actions and reduce multicollinearity for regression use was also difficult. Discussions with professors helped narrow down the scope for the former, whereas the latter was navigated by revisiting the basics learned in the program. In terms of successes, the project effectively conveys the story of Argentina's campaign in a digestible manner while also providing actionable insights. As for shortcomings, certain goals from the project proposal — such as synergy graphs to highlight dynamic player relationships, and machine learning models for prediction of win probability — could not be executed due to time constraints. However, these limitations did not significantly detract from the overall aims and outcomes of the project.

There are plans to continue working on this project after completion of the MS program. Given the success of the initial version, there is an interest in expanding it to include a win-probability model trained on data from previous World Cups. This would enable analysis of how in-game management actions such as substitutions and formation adjustments affect the evolving probability of winning a match. The project will also be shared with football analysts and the broader football community via social media to invite feedback. Nevertheless, the project's archival structure allows for future expansion should additional time or opportunities arise.

APPENDIX

Data

All data files are listed here and can be found in the main Github repository:
https://github.com/a-partha/Masters-Capstone-Project

1) Comprehensive Tournament Data:

    a) fifa_2022_matches.json
    Full list of 2022 World Cup game data for all teams.

    b) argentina_stats.csv
    Team-level defensive and offensive statistics for Argentina during the 2022 tournament.

    c) non_argentina_worldcup_events.jsonl.gz
    Compressed JSON Lines file containing event-level data for all 2022 games excluding Argentina.

2) Shots and Pressure Data:

    a) arg_shots_summary.csv
    Aggregated shot information including xG and positional metadata for Argentina's games.

    b) standard_ppda.csv
    Argentina's defensive pressure intensity metric for each game for press analysis comparisons.

    c) pressure_files
    Folder with processed JSON files for Argentina's under_pressure events per game.

Data Dictionary

Key Variable Name	Data Type	Description & Role in Analysis
df_argentina	table	Master match-level summary for Argentina's 2022 World Cup games. Drives iteration logic and links event files to opponents and game contexts.

df_ppda	table	Aggregated data used in regression, combining pressing intensity (PPDA) with key defensive actions such as tackles, fouls committed, and interceptions.
df_pressure	table	Pressure events by Argentina in the HPZ during regular time (0’ to 90’). Used for heatmaps and temporal analyses.
df_shots	table	Shot-level event data including shot location, outcome, and time. Used for timelines and to compute pressure-linked attacking output.
statsbomb_xg	float	Expected goals (xG) values directly derived from Hudl StatsBomb for Argentina’s shots. Used to evaluate finishing efficiency.
xt_grid	array	Expected threat value matrix, trained on non-Argentina data. Serves as the neutral baseline to evaluate Argentina’s ball progression value.
build_up_sequences	list	Structured sequences involving >=3 passes or progressive carries. Used to assess how Argentina advanced play into dangerous attacking zones.
tooltips[opponent]	dictionary	Annotation dictionary per opponent containing key events (goals, substitutions, formation shifts) for visual timelines.
pressure_series	time series	Minute-wise count of under_pressure events performed by Argentina during each match. Used to visualize how frequently they retained composure or played through pressure over time in in-game management plots.

Glossary of Functions

Function Name	File	Purpose
animateDots	logo-dots.js	Makes the tournament logo's dots appear and siappear as the reader scrolls.
createArcPath	pitch.js, pitch_context.js	Creates a ready-made D3 arc so the penalty-area curves are easy to draw.
initScroll	scroll.js	Starts the main Scrollama observers and keeps them responsive.
setupScrollStack	scroll2.js	Runs a second Scrollama that reveals stacked text once the chart is in view.
initializeHeatmap	heatmap.js	Initializes the heat-map grid, color scale, and first render after the pitch loads.
updateHeatmap	heatmap.js	Loads the current game's pressure data and redraws the heatmap overlay.
update	igm.js	Refreshes the line-plot for the selected opposition.

REFERENCES

[1] StatsBomb, “Free Data.” [Online]. Available: https://statsbomb.com/what-we-do/hub/free-data/

[2] D. Pleuler, “Soccer Analytics Handbook.” GitHub. [Online]. Available: https://github.com/devinpleuler/analytics-handbook/blob/master/soccer_analytics_handbook.ipynb

[3] StatsBomb, “Open Data Specification v1.1.” GitHub. [Online]. Available: https://github.com/statsbomb/open-data/blob/master/doc/StatsBomb%20Open%20Data%20Specification%20v1.1.pdf

[4] L. Lolli, P. Bauer, C. Irving, D. Bonanno, O. Höner, W. Gregson, and V. Di Salvo, "Data analytics in the football industry: A survey investigating operational frameworks and practices in professional clubs and national federations from around the world," *Science and Medicine in Football*, vol. 9, no. 2, pp. 1–10, May 2024. [Online]. Available: https://doi.org/10.1080/24733938.2024.2341837

[5] M. Carey, “How data departments have evolved and spread across English football clubs,” *The New York Times*, Sep. 3, 2024. [Online]. Available: https://www.nytimes.com/athletic/5697684/2024/09/03/football-analytics-uk-evolution/

[6] D. Burdeno, “Exploring Data Analytics in Soccer,” *Flatiron School Blog*, Feb. 26, 2024. [Online]. Available: https://flatironschool.com/blog/exploring-data-analytics-in-soccer/

[7] M. Carey, “The genie is well and truly out of the bottle when it comes to data in football,” *The New York Times*, Mar. 26, 2022. [Online]. Available: https://www.nytimes.com/athletic/3209373/2022/03/26/the-genie-is-well-and-truly-out-of-the-bottle-when-it-comes-to-data-in-football/

[8] The Athletic FC, *YouTube* Channel. [Online]. Available: https://www.youtube.com/channel/UC6ZMmQaL9wZYo4iLw8n8xiA

[9] P. Power, H. Ruiz, X. Wei, and P. Lucey, "Not all passes are created equal: Objectively measuring the risk and reward of passes in soccer from tracking data," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, Halifax, NS, Canada: ACM, 2017, pp. 1605–1613. [Online]. Available: https://doi.org/10.1145/3097983.3098051

[10] S. Rudd, “A framework for tactical analysis and individual offensive production assessment in soccer using Markov Chains,” [Online]. Available: https://nessis.org/nessis11/rudd.pdf

[11] K. Singh, “Introducing Expected Threat (xT),” *Karun.in* Blog. [Online]. Available: https://karun.in/blog/expected-threat.html

[12] Football Made Simple, *YouTube* Channel. [Online]. Available: https://www.youtube.com/channel/UCFY0YHhxiIQWYYsLgeUBcbg

[13] M. Bostock and Observable, Inc., *D3.js: Data-Driven Documents*. [Online]. Available: https://d3js.org/

[14] GitHub, Inc., *GitHub Pages*. [Online]. Available: https://pages.github.com/

[15] R. Samora, *Scrollama: A JavaScript library for scrollytelling*. [Online]. Available: https://github.com/russellsamora/scrollama

[16] Google, *Google Colab*. [Online]. Available: https://colab.research.google.com/

[17] Anysphere Inc., *Cursor IDE*. [Online]. Available: https://www.cursor.com/

[18] A. Durgapal and A. Rowlinson, *mplsoccer: Python library for football analytics*. GitHub. [Online]. Available: https://github.com/andrewRowlinson/mplsoccer